

# Non-asymptotic Confidence Sets for Extrinsic Means on Spheres and Projective Spaces

Thomas Hotz, Florian Kelma

Institut für Mathematik  
Technische Universität Ilmenau

## Abstract

Confidence sets from i.i.d. data are constructed for the extrinsic mean of a probability measure  $\mathbf{P}$  on spheres, real projective spaces  $\mathbb{RP}^k$ , and complex projective spaces  $\mathbb{CP}^k$ , as well as Grassmann manifolds, with the latter three embedded by the Veronese-Whitney embedding. When the data are sufficiently concentrated, these are projections of a ball around the corresponding Euclidean sample mean. Furthermore, these confidence sets are rate-optimal. The usefulness of this approach is illustrated for projective shape data.

## 1 Introduction and definitions

Data assuming values on unit spheres, particularly on the circle, as well as on real projective spaces arise frequently in applications, examples being measurements of wind directions or axial data describing optical axes of crystals (Mardia and Jupp; 2000). Moreover, we consider data assuming values on real and complex Grassmann manifolds which play important roles in several shape spaces which arise in image-processing applications, cf. (Kendall et al.; 1999; Patrangenaru and Mardia; 2003; Mardia and Patrangenaru; 2005).

Motivated by these applications, we will concern ourselves with the arguably simplest statistic, the mean. However, since there is neither a natural addition of points on these manifolds nor the division by a natural number, the meaning of “mean” is a priori unclear.

We assume the data can be modelled as independent random variables  $Z_1, \dots, Z_n$  which are identically distributed as the random variable  $Z$  taking values in a closed subset  $M$  of the unit sphere  $S^k = \{x \in \mathbb{R}^{k+1} : \|x\| = 1\}$ . Using the Veronese-Whitney embedding, real and complex Grassmann manifolds  $\text{Gr}(m, \mathbb{R}^{d+1})$  resp.  $\text{Gr}(m, \mathbb{C}^{d+1})$ ,  $m < d+1$ , can be seen as closed subspaces of unit spheres allowing to treat these cases, too. Recall that  $\text{Gr}(1, \mathbb{R}^{d+1}) = \mathbb{RP}^d$ , and analogously  $\text{Gr}(1, \mathbb{C}^{d+1}) = \mathbb{CP}^d$  are the real and complex projective spaces of dimension  $d$ .

Since any unit sphere is a subset of a Euclidean vector space, the Euclidean sample mean  $\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i$  is well-defined, but  $\bar{Z}_n$  cannot be taken as a mean of the sphere since it may have norm less than 1. Though, the Euclidean sample mean is the minimiser of the sum of squared distances, and thus this can be put in the more general framework of Fréchet means, see (Fréchet; 1948): define the set of extrinsic sample means to be

$$\hat{\mu}_n = \underset{\nu \in M}{\operatorname{argmin}} \sum_{i=1}^n \|Z_i - \nu\|^2, \quad (1)$$

and analogously define the set of extrinsic population means of the random variable  $Z$  to be

$$\mu = \underset{\nu \in M}{\operatorname{argmin}} \mathbf{E} \|Z - \nu\|^2$$

where  $M \subseteq S^k$  is the subset of the possible values of  $Z$ . As usual, the extrinsic sample means are the extrinsic population means when considering the empirical distribution of  $Z_1, \dots, Z_n$ .

The extrinsic population mean is closely related to the Euclidean population mean  $\mathbf{E}Z$ :

$$\mu = \underset{\nu \in M}{\operatorname{argmin}} \|\mathbf{E}Z - \nu\|^2$$

since  $\mathbf{E} \|Z - \nu\|^2 = \mathbf{E} \|Z - \mathbf{E}Z\|^2 + \|\mathbf{E}Z - \nu\|^2$ . Hence,  $\mu$  is the set of points on the manifold closest to  $\mathbf{E}Z$ , and  $\mu$  is unique if and only if the orthogonal projection of  $\mathbf{E}Z$  onto  $M$  is unique;

for  $M = S^k$ , this is the case if and only if  $\mathbf{E}Z \neq 0$ , since the orthogonal projection of  $\mathbf{E}Z$  is then given by  $\mu = \mathbf{E}Z/\|\mathbf{E}Z\|$ ; otherwise, i.e. if  $\mathbf{E}Z = 0$ , the set of extrinsic population means is all of  $S^k$ . Analogous results hold for the extrinsic sample mean.

In this article, we aim to construct non-asymptotic confidence sets for the extrinsic population mean without making any assumptions about the distribution of the data besides the assumption that the data are independent and identically distributed. It has been shown in (Hotz; 2013) and (Hotz et al.; 2015) that this is possible for data on the circle. We will generalize and improve the formalism used in (Hotz; 2013) to obtain confidence sets in this more general case.

First, we construct confidence sets for the Euclidean population mean noting that the projection of these sets are confidence sets for the extrinsic population mean. The constructed confidence sets for the Euclidean population mean are no open balls, but we will see (Corollary 3) that a  $(1 - \alpha)$ -confidence set for the extrinsic population mean is given by the projection of the open ball around the Euclidean sample mean  $\bar{Z}_n$  with radius

$$\varepsilon = \sqrt{\frac{1}{\alpha n} \left(1 - \|\bar{Z}_n\|^2 + \frac{1}{\alpha n}\right)}.$$

So, the remaining problem for arbitrary closed  $M \subseteq S^k$  is to understand the projection of open balls.

In the simple case where  $M = S^k$ , we immediately obtain

$$\left\{ x \in S^k : \sin \angle(x, \hat{\mu}_n) < \sqrt{\frac{1 - \|\bar{Z}_n\|^2 + \frac{1}{\alpha n}}{\alpha n \|\bar{Z}_n\|^2}} \right\}$$

as a  $(1 - \alpha)$ -confidence set, if  $\|\bar{Z}_n\|^2 \geq \frac{1}{\alpha n}$  (Proposition 6). We further note that this condition is violated with exponentially decreasing probability for  $\mathbf{E}Z \neq 0$ , and fulfilled with probability at most  $\alpha$  for  $\mathbf{E}Z = 0$ .

Additionally, we cover complex Grassmann manifolds  $\text{Gr}(m, \mathbb{C}^{k+1})$  in Section 4. These spaces can be considered as closed subspaces of a  $((k+1)^2 - 1)$ -dimensional sphere by the *Veronese-Whitney embedding* into the Euclidean space of Hermitian matrices with the Euclidean norm  $\|\cdot\|$  being called Frobenius norm  $\|\cdot\|_F$  there. In this case, the projection  $\pi$  maps a Hermitian matrix to the span of eigenvectors corresponding to its  $m$  largest eigenvalues. Using basic results from linear algebra, we obtain an open  $(1 - \alpha)$ -confidence ball around the extrinsic sample mean  $\hat{\mu}_n$  with radius

$$\delta_n = \frac{\sqrt{2} \sqrt{1 - \|\bar{Z}_n\|_F^2 + \frac{1}{\alpha n}}}{\sqrt{\alpha n}(\hat{\lambda}_m - \hat{\lambda}_{m+1}) - \sqrt{1 - \|\bar{Z}_n\|_F^2 + \frac{1}{\alpha n}}}$$

if the  $m$ -th largest and  $m+1$ -st largest eigenvalues  $\hat{\lambda}_m$  and  $\hat{\lambda}_{m+1}$  of the Euclidean sample mean  $\bar{Z}_n$  are sufficiently separated (Proposition 12). Again, this condition is violated with exponentially decreasing probability for uniquely projected  $\mathbf{E}Z$ , and fulfilled with probability at most  $\alpha$  for non-uniquely projected  $\mathbf{E}Z$ .

This can then be applied to shape data, examples of which are shown in Section 5. We conclude with a discussion of the results.

## 2 Confidence sets for Euclidean and extrinsic means

As above, let  $Z_1, \dots, Z_n$ ,  $n \in \mathbb{N}$  be i.i.d. random elements on  $M$  where  $M$  is a closed subset  $M$  of the unit sphere  $S^k = \{x \in \mathbb{R}^{k+1} : |x| = 1\}$ . Further, let  $\pi : \mathbb{R}^{k+1} \rightarrow M$  be the projection of  $\mathbb{R}^{k+1}$  to  $M$  in the sense of best approximation. With this setup, we have the following multivariate version of Chebyshev's inequality:

**Theorem 1** For  $\alpha \in (0, 1)$ ,

$$\mathbf{P} \left( \|\bar{Z}_n - \mathbf{E}Z\|^2 \geq \frac{1 - \|\mathbf{E}Z\|^2}{\alpha n} \right) \leq \alpha.$$

PROOF. Observe that  $\mathbf{E} \|Z - \mathbf{E}Z\|^2 = 1 - \|\mathbf{E}Z\|^2$ , such that by the assumption of observing i.i.d. data,  $\mathbf{E} \|\bar{Z}_n - \mathbf{E}Z\|^2 = \frac{1}{n}(1 - \|\mathbf{E}Z\|^2)$ . Now apply Markov's inequality.

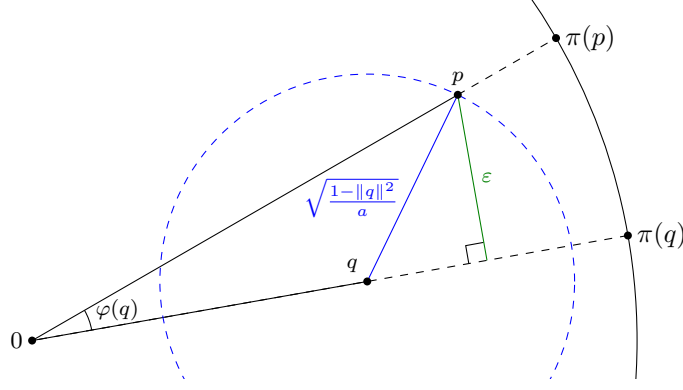


Figure 1: Sketch for the proof of **Proposition 2**.

This simple inequality gives a level  $\alpha$  test of the hypothesis  $\mathbf{E}Z = q$  for any fixed  $q$  with  $\|q\| \leq 1$  (note that necessarily  $\|\mathbf{E}Z\| \leq 1$  by convexity): simply reject  $q$  if  $\|\bar{Z}_n - q\|^2 \geq \frac{1 - \|q\|^2}{\alpha n}$ . The acceptance set of that test,

$$C(p, a) := \left\{ q : \|q\| \leq 1 \wedge \|p - q\|^2 < \frac{1 - \|q\|^2}{a} \right\}$$

with  $p = \bar{Z}_n$  and  $a = \alpha n$ , is therefore a  $(1 - \alpha)$ -confidence set for the Euclidean population mean. Henceforth, the projection of this set under  $\pi$  is a  $(1 - \alpha)$ -confidence set for the extrinsic population mean.

**Proposition 2** For  $r > 0$ , define the open ball

$$B_r(p) := \{q \in E : \|p - q\| < r\}$$

around  $p$  with radius  $r$ . Then

$$\pi(C(p, a)) = \pi\left(B_{\sqrt{\frac{1}{a}(1 - \|p\|^2 + \frac{1}{a})}}(p)\right).$$

In particular, if  $\|p\|^2 \geq \frac{1}{a}$ , then  $C(p, a)$  is contained in the positive cone over

$$B_{\sqrt{\frac{1}{a}(1 - \|p\|^2 + \frac{1}{a})}}(p).$$

PROOF. We will bound the angle  $\varphi$  between  $p$  and arbitrary  $q \in C(p, a)$ . The angle is given by

$$\varphi = \arccos \frac{\langle p, q \rangle}{\|p\| \|q\|},$$

hence we will compute the minimum of  $\cos \varphi = \frac{\langle p, q \rangle}{\|p\| \|q\|}$  with respect to  $q$  under the constraint

$$\begin{aligned} \|p - q\|^2 &= \frac{1 - \|q\|^2}{a} \\ \Leftrightarrow \|p\|^2 - 2\langle p, q \rangle + \|q\|^2 &= \frac{1 - \|q\|^2}{a} \\ \Leftrightarrow \langle p, q \rangle &= \frac{1}{2} \left( \|p\|^2 + \|q\|^2 - \frac{1 - \|q\|^2}{a} \right), \end{aligned}$$

see **Figure 1**. Hence,

$$\frac{\|p\|^2 + \|q\|^2 - \frac{1 - \|q\|^2}{a}}{2\|p\| \|q\|} = \frac{\|p\|^2 + \frac{a+1}{a}\|q\|^2 - \frac{1}{a}}{2\|p\| \|q\|}$$

has to be minimized with respect to  $\|q\|$ . Starting from

$$0 = \frac{\partial}{\partial \|q\|} \left( \frac{\|p\|^2 + \frac{a+1}{a}\|q\|^2 - \frac{1}{a}}{2\|p\| \|q\|} \right) \Bigg|_{q_{max}} = \frac{1}{2\|p\|} \frac{2\frac{a+1}{a}\|q_{max}\|^2 - \|p\|^2 - \frac{a+1}{a}\|q_{max}\|^2 + \frac{1}{a}}{\|q_{max}\|^2},$$

one obtains

$$\|q_{max}\| = \sqrt{\frac{a}{a+1} \left( \|p\|^2 - \frac{1}{a} \right)},$$

and thus

$$\begin{aligned} \varepsilon_{max}^2 &= \|p\|^2 \sin^2 \varphi_{max} \\ &= \|p\|^2 (1 - \cos^2 \varphi_{max}) \\ &= \frac{1}{4\|q_{max}\|^2} \left( 4\|p\|^2 \|q_{max}\|^2 - \left( \|p\|^2 + \frac{a+1}{a} \|q_{max}\|^2 - \frac{1}{a} \right)^2 \right) \\ &= \frac{a+1}{4a(\|p\|^2 - \frac{1}{a})} \left( 4\frac{a}{a+1} \|p\|^2 \left( \|p\|^2 - \frac{1}{a} \right) - 4 \left( \|p\|^2 - \frac{1}{a} \right)^2 \right) \\ &= \frac{a+1}{a(\|p\|^2 - \frac{1}{a})} \left( \|p\|^2 - \frac{1}{a} \right) \left( \frac{a}{a+1} \|p\|^2 - \|p\|^2 + \frac{1}{a} \right) \\ &= -\frac{1}{a} \|p\|^2 + \frac{a+1}{a^2} \\ &= \frac{1}{a} \left( 1 - \|p\|^2 + \frac{1}{a} \right) \end{aligned}$$

which finishes the proof.  $\square$

**Corollary 3** *A  $(1 - \alpha)$ -confidence set for the extrinsic population mean is given by the projection of the ball around  $\bar{Z}_n$  with radius*

$$\varepsilon = \sqrt{\frac{1}{\alpha n} \left( 1 - \|\bar{Z}_n\|^2 + \frac{1}{\alpha n} \right)}.$$

*More precisely, the probability that  $\pi(B_\varepsilon(\bar{Z}_n))$  covers the extrinsic mean set is at least  $1 - \alpha$ .*

*In addition,  $0 \in B_\varepsilon(\bar{Z}_n)$  if and only if  $\|\bar{Z}_n\|^2 < \frac{1}{\alpha n}$ . In that case,  $\pi(B_\varepsilon(\bar{Z}_n)) = M$ .*

**Remark. 4** *As an alternative to the use of Chebychev's inequality, a multivariate version of Hoeffding's inequality could be used to get confidence sets. After some calculations following the lines of (Boucheron et al.; 2013, Ex. 6.3), one can obtain*

$$\mathbf{P} \left( \|\bar{Z}_n - \mathbf{EZ}\| > \frac{1}{\sqrt{n}} \left( \sqrt{-2(1 + \|\mathbf{EZ}\|^2 \log \alpha)} + \sqrt{1 - \|\mathbf{EZ}\|^2} \right) \right) \leq \alpha.$$

*For  $\alpha = .05$  and  $\mathbf{EZ} = 0$ , this radius is roughly 23% smaller than the corresponding radius obtained by Chebyshev's inequality. However, the use of Chebyshev's inequality gives smaller confidence sets for  $\|\mathbf{EZ}\|$  close to 1 which is important since directional data are often highly concentrated.*

### 3 Confidence sets for directional data

In this section, let  $M = S^k := \{x \in \mathbb{R}^{k+1} : \|x\| = 1\}$  be the entire  $k$ -dimensional sphere. Clearly,  $\pi$  is now given by

$$\pi(x) = \operatorname{argmin}_{\nu \in S^k} \|\nu - x\| = \frac{x}{\|x\|}$$

for  $x \neq 0$ , while  $\pi(0) = S^k$ . The following statement is well-known; see e.g. (Bhattacharya and Patrangenaru; 2003) or (Mardia and Jupp; 2000).

**Theorem 5** *The extrinsic population mean  $\mu$  of  $\mathbf{P}^Z$  is unique if and only if  $\mathbf{EZ} \neq 0$ . In this case,  $\mu = \frac{\mathbf{EZ}}{\|\mathbf{EZ}\|}$ , otherwise  $\mu = S^k$ . Similarly, the extrinsic sample mean  $\hat{\mu}_n$  is unique if and only if  $\bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i \neq 0$ , whence  $\hat{\mu}_n = \frac{\bar{Z}_n}{\|\bar{Z}_n\|}$ , otherwise  $\hat{\mu}_n = S^k$ .*

Using Corollary 3, we immediately obtain confidence sets for the extrinsic population mean.

**Proposition 6** *Let  $\alpha \in (0, 1)$ . Then  $\pi(C(\bar{Z}_n, \alpha n))$  is a  $(1 - \alpha)$ -confidence set for the extrinsic population mean (set) of directional data.*

(i) If  $\|\bar{Z}_n\|^2 \geq \frac{1}{\alpha n}$ , then  $\pi(C(\bar{Z}_n, \alpha n))$  is given by

$$\left\{ x \in S^k : \sin \angle(x, \hat{\mu}_n) < \sqrt{\frac{1 - \|\bar{Z}_n\|^2 + \frac{1}{\alpha n}}{\alpha n \|\bar{Z}_n\|^2}} \right\},$$

which, in the metric on  $S^k$  given by arc length, is the open ball around  $\hat{\mu}_n$  with radius

$$\arcsin \sqrt{\frac{1 - \|\bar{Z}_n\|^2 + \frac{1}{\alpha n}}{\alpha n \|\bar{Z}_n\|^2}}.$$

Otherwise,  $\pi(C(\bar{Z}_n, \alpha n)) = S^k$ , the closed ball around  $\hat{\mu}_n$  with radius  $\pi$ .

(ii) If  $\mathbf{E}Z = 0$ , i.e. if the extrinsic population mean set equals  $S^k$ , then  $\|\bar{Z}_n\|^2 \geq \frac{1}{\alpha n}$  with probability at most  $\alpha$ , so indeed  $\pi(C(\bar{Z}_n, \alpha n)) = S^k$  with probability at least  $(1 - \alpha)$ .

(iii) If  $\mathbf{E}Z \neq 0$ , i.e. if the extrinsic population mean is unique, then

$$\mathbf{P}(\|\bar{Z}_n\| \leq \frac{1}{\alpha n}) \leq e^{-\frac{n}{2}(\|\mathbf{E}Z\| - \frac{1}{\alpha n})^2},$$

hence the probability of obtaining the trivial confidence set  $\pi(C(\bar{Z}_n, \alpha n)) = S^k$  goes to zero at an exponential rate when  $n$  tends to infinity. Furthermore, the radius of the confidence ball in (i) is in  $O_{\mathbf{P}}(\frac{1}{\sqrt{n}})$ .

PROOF. For (i), see [Corollary 3](#), and note that the projection  $\pi$  preserves the angle between two points. If  $\|\bar{Z}_n\|^2 \geq \frac{1}{\alpha n}$ , then the boundary for the angle between a point in the  $(1 - \alpha)$ -confidence set for the Euclidean mean and the Euclidean sample mean is given by

$$\sin \varphi_{\max} = \frac{\sqrt{\frac{1}{\alpha n} (1 - \|\bar{Z}_n\|^2 + \frac{1}{\alpha n})}}{\|\bar{Z}_n\|}.$$

Hence, this is the boundary for the angle after projecting to  $S^k$ . If  $\|\bar{Z}_n\|^2 < \frac{1}{\alpha n}$ , then  $0 \in C(\bar{Z}_n, \alpha n)$  whence  $\pi(C(\bar{Z}_n, \alpha n)) = S^k$ .

(ii) follows directly from [Theorem 1](#).

For (iii), we use Hoeffding's inequality, cf. ([Hoeffding; 1963](#)):

$$\begin{aligned} \mathbf{P}(\|\bar{Z}_n\| \leq \frac{1}{\alpha n}) &\leq \mathbf{P}(|\langle \bar{Z}_n, \frac{\mathbf{E}Z}{\|\mathbf{E}Z\|} \rangle| \leq \frac{1}{\alpha n}) \\ &\leq \mathbf{P}(|\langle \bar{Z}_n, \frac{\mathbf{E}Z}{\|\mathbf{E}Z\|} \rangle - \|\mathbf{E}Z\|| \geq \|\mathbf{E}Z\| - \frac{1}{\alpha n}) \\ &\leq e^{-\frac{n}{2}(\|\mathbf{E}Z\| - \frac{1}{\alpha n})^2} \end{aligned}$$

Then, the second statement of (iii) follows directly from (i).  $\square$

For data on the circle  $S^1$  however, we suggest to construct confidence sets using Hoeffding's inequality as in ([Hotz et al.; 2015](#)).

**Remark. 7** Using the central limit theorem for the extrinsic mean, cf. ([Jammalamadaka and SenGupta; 2001](#)), one can obtain asymptotic confidence sets

$$\left\{ x \in S^k : \angle(x, \hat{\mu}_n) < q_{1-\frac{\alpha}{2}} \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n \sin^2 \angle(Z_i, \bar{Z}_n)}{n \|\bar{Z}_n\|^2}} \right\}$$

where  $q_{1-\frac{\alpha}{2}}$  is the  $(1 - \frac{\alpha}{2})$ -quantil of the standard normal distribution, but these do not guarantee coverage for finite  $n$  which is problematic even for considerably large sample sizes, cf. ([Hotz et al.; 2015](#)).

## 4 Confidence sets for data on Grassmann manifolds

In this section, let  $\text{Gr}(m, \mathbb{C}^{d+1})$ ,  $m < d+1$ , be the Grassmann manifold of  $m$ -dimensional complex subspaces of  $\mathbb{C}^{d+1}$ . Analogous results hold for real Grassmann manifolds. The projective spaces  $\mathbb{CP}^d = \text{Gr}(1, \mathbb{C}^{d+1})$  and  $\mathbb{RP}^d = \text{Gr}(1, \mathbb{R}^{d+1})$  are special cases of Grassmann manifolds<sup>1</sup>.

We embed the Grassmann manifold  $\text{Gr}(m, \mathbb{C}^{d+1})$  into the  $(d+1)^2$ -dimensional Euclidean space  $(\text{Herm}(d+1), \text{tr}(\cdot, \cdot))$  of Hermitian matrices by choosing an orthonormal basis  $\{u_1, \dots, u_m\}$  for a given subspace  $U \in \text{Gr}(m, \mathbb{C}^{d+1})$  and mapping it to

$$\iota(U) = \frac{1}{\sqrt{m}} \sum_{i=1}^m u_i u_i^*$$

which is independent of the choice of the basis, making the mapping well-defined;  $\iota : \text{Gr}(m, \mathbb{C}^{d+1}) \rightarrow \text{Herm}(d+1)$  is called Veronese-Whitney embedding. As a mapping,  $\iota(U)$  is actually the projection from  $\mathbb{C}^{k+1}$  onto the subspace  $U$ , so  $\iota$  is injective, allowing us to identify  $\text{Gr}(m, \mathbb{C}^{d+1})$  with  $M = \iota(\text{Gr}(m, \mathbb{C}^{d+1}))$ . Note that [Corollary 3](#) holds for data in  $\text{Gr}(m, \mathbb{C}^{k+1})$  since  $\|\iota(U)\|_F = \sqrt{\text{tr}(\iota(U)\iota(U)^*)} = 1$  for all  $U \in \text{Gr}(m, \mathbb{C}^{d+1})$  with  $\|\cdot\|_F$  being the Frobenius norm, i.e. the standard Euclidean norm on  $\mathbb{C}^{(d+1) \times (d+1)}$  viewed as  $\mathbb{C}^{(d+1)^2}$ . Therefore,  $M \subseteq S^k$  for  $k = (d+1)^2 - 1$ , allowing us to apply the results of [Section 2](#).

The Frobenius norm then equips  $\text{Gr}(m, \mathbb{C}^{k+1})$  with an extrinsic metric  $d$  defined by

$$\begin{aligned} d^2(U, V) &= \|\iota(U) - \iota(V)\|_F^2 \\ &= \frac{1}{m} \left( \sum_{i=1}^m \text{tr}(u_i u_i^*) + \sum_{i=1}^m \text{tr}(v_i v_i^*) - 2 \sum_{i,j=1}^m \text{tr}(u_i u_i^* v_j v_j^*) \right) \\ &= 2 \left( 1 - \frac{1}{m} \sum_{i,j=1}^m |v_j^* u_i|^2 \right) \\ &= 2(1 - \|\iota(U)\iota(V)\|_F^2). \end{aligned}$$

As always,  $\pi : \text{Herm}(k+1) \rightarrow M$  shall be the projection in the sense of best approximation. Here,  $\pi$  maps a matrix to the span of  $m$  linear independent eigenvectors to its largest eigenvalues. The image of matrix under  $\pi$  is unique if and only if the direct sum of the eigenspaces corresponding to the  $m$  largest eigenvalues (counted with multiplicities) is  $m$ -dimensional.

**Proposition 8** *Let  $A$  be a Hermitian matrix with eigenvalues  $\lambda_1 \geq \dots \geq \lambda_m > \lambda_{m+1} \geq \dots \geq \lambda_{k+1}$ , so that  $\pi(A)$  is unique. If  $\lambda_m - \lambda_{m+1} \geq \sqrt{2}\varepsilon > 0$ , then  $\pi$  is unique on  $B_\varepsilon(A) \subset \text{Herm}(k+1)$ .*

For the proof, we need the theorem of Wielandt-Hoffman, cf. ([Wielandt and Hoffman; 1953](#)):

**Theorem 9 (Wielandt-Hoffman)** *If  $A$  and  $A + E$  are normal matrices with eigenvalues  $\lambda_1 \geq \dots \geq \lambda_{k+1}$  respectively  $\sigma_1 \geq \dots \geq \sigma_{k+1}$ , then*

$$\sum_{j=1}^{k+1} (\lambda_i - \sigma_i)^2 \leq \|E\|_F^2.$$

**PROOF OF PROPOSITION 8.** Let  $\sigma_1 \geq \dots \geq \sigma_{k+1}$  be the eigenvalues of  $A + E \in B_\varepsilon(A)$ . It has to be shown that  $\sigma_m - \sigma_{m+1} > 0$  for all  $A + E \in B_\varepsilon(A)$ , i.e. for all  $E \in \text{Herm}(k+1)$  with  $\|E\|_F < \varepsilon$ .

Suppose that  $\sigma_m = \sigma_{m+1}$ . Then

$$(\lambda_m - \sigma_m)^2 + (\lambda_{m+1} - \sigma_{m+1})^2$$

is minimal for  $\sigma_m = \sigma_{m+1} = \frac{\lambda_m + \lambda_{m+1}}{2}$ . Hence,

$$\|E\|_F^2 = \sum_{j=1}^{k+1} (\lambda_i - \sigma_i)^2 \geq (\lambda_m - \sigma_m)^2 + (\lambda_{m+1} - \sigma_{m+1})^2 \geq \varepsilon^2,$$

but that contradicts the [Wielandt-Hoffman Theorem](#) since  $\|E\|_F < \varepsilon$ .  $\square$

To bound the projection of the ball in [Corollary 3](#), we use the Davis-Kahan  $\sin \theta$  Theorem (cf. ([Davis and Kahan; 1969](#))), and conclude the following statements:

<sup>1</sup>Note that  $\mathbb{RP}^1 \cong S^1$  and  $\mathbb{CP}^1 \cong S^2$ . We suggest using the results of [Section 3](#) to obtain confidence sets from data in these spaces since a lower dimensional Euclidean space is used there.

**Theorem 10** Let  $A$  and  $A + E$  be Hermitian matrices with eigenvalues  $\lambda_1 \geq \dots \geq \lambda_m > \lambda_{m+1} \geq \dots \geq \lambda_{k+1}$  resp.  $\sigma_1 \geq \dots \geq \sigma_m > \sigma_{m+1} \geq \dots \geq \sigma_{k+1}$  with  $\sigma_m > \lambda_{m+1}$ . Then,

$$(\sigma_m - \lambda_{m+1}) \cdot d(\pi(A), \pi(A + E)) \leq \sqrt{2} \|E\|_F.$$

**Proposition 11** Let  $A$  be a Hermitian matrix with eigenvalues  $\lambda_1 \geq \dots \geq \lambda_m > \lambda_{m+1} \geq \dots \geq \lambda_{k+1}$ . If  $\lambda_m - \lambda_{m+1} \geq \sqrt{2}\varepsilon > 0$ , then

$$\pi(B_\varepsilon(A)) \subseteq B_\delta(\pi(A))$$

in  $\text{Gr}(m, \mathbb{C}^{k+1})$  where  $\delta = \frac{\sqrt{2}\varepsilon}{\lambda_m - \lambda_{m+1} - \varepsilon}$ .

PROOF. First, note that  $B_\varepsilon(A)$  is uniquely projected by  $\pi$  due to **Proposition 8** since  $\lambda_m - \lambda_{m+1} \geq \sqrt{2}\varepsilon$ . Let  $E$  be an arbitrary Hermitian matrix with  $\|E\|_F < \varepsilon$ , i.e.  $A + E \in B_\varepsilon(A)$ , and denote the eigenvalues of  $A + E$  by  $\sigma_1 \geq \dots \geq \sigma_{k+1}$ . By construction and the Wielandt-Hoffman Theorem,

$$\sigma_m - \lambda_j = \underbrace{\sigma_m - \lambda_m}_{> -\varepsilon} + \lambda_m - \lambda_j > \lambda_m - \lambda_{m+1} - \varepsilon$$

for all  $j > m$ . Using **Theorem 10**,

$$d(\pi(A), \pi(A + E)) \leq \frac{\sqrt{2} \|E\|_F}{\sigma_m - \lambda_{m+1}} < \frac{\sqrt{2}\varepsilon}{\lambda_m - \lambda_{m+1} - \varepsilon},$$

which is what had to be shown.  $\square$

Using these results and **Corollary 3**, we obtain confidence sets for the extrinsic population mean on  $\text{Gr}(m, \mathbb{C}^{d+1})$ .

**Proposition 12** Let  $\alpha \in (0, 1)$ , and suppose  $\bar{Z}_n$  to have eigenvalues  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_{k+1}$ .

(i) If

$$\hat{\lambda}_m - \hat{\lambda}_{m+1} \geq \sqrt{\frac{2(1 - \|\bar{Z}_n\|_F^2 + \frac{1}{\alpha n})}{\alpha n}},$$

then  $\pi(C(\bar{Z}_n, \alpha n))$  is given by  $D = B_{\delta_n}(\hat{\mu}_n) \subset \text{Gr}(m, \mathbb{C}^{k+1})$  where

$$\delta_n = \frac{\sqrt{2}\sqrt{1 - \|\bar{Z}_n\|_F^2 + \frac{1}{\alpha n}}}{\sqrt{\alpha n}(\hat{\lambda}_m - \hat{\lambda}_{m+1}) - \sqrt{1 - \|\bar{Z}_n\|_F^2 + \frac{1}{\alpha n}}}.$$

Choosing  $D = M$ , the closed ball around  $\hat{\mu}_n$  with radius  $\sqrt{2}$ , otherwise ensures that the extrinsic mean  $\mu$  is covered by  $D$  with probability at least  $(1 - \alpha)$ .

(ii) If  $\mathbf{EZ}$  is not uniquely projected, then

$$\hat{\lambda}_m - \hat{\lambda}_{m+1} < \sqrt{\frac{2(1 - \|\bar{Z}_n\|_F^2 + \frac{1}{\alpha n})}{\alpha n}}$$

with probability at least  $(1 - \alpha)$ .

(iii) If there is a unique extrinsic population mean, then

$$\mathbf{P} \left( \hat{\lambda}_m - \hat{\lambda}_{m+1} < \sqrt{\frac{2(1 - \|\bar{Z}_n\|_F^2 + \frac{1}{\alpha n})}{\alpha n}} \right) \leq e^{-\frac{1}{2} \left( \sqrt{n} \frac{\sigma_m - \sigma_{m+1}}{\sqrt{2}} - \sqrt{\frac{1 + \frac{1}{\alpha n}}{\alpha}} - \sqrt{1 - \|\mathbf{EZ}\|} \right)^2} \xrightarrow{n \rightarrow \infty} 0$$

where  $\sigma_m, \sigma_{m+1}$  are the  $m$ -th and  $(m+1)$ -th eigenvalues of  $\mathbf{EZ}$ . In particular, the probability of obtaining the trivial confidence set goes to zero at an exponential rate when  $n$  tends to infinity. Furthermore, the radius of the confidence ball in (i) is in  $O_{\mathbf{P}}(\frac{1}{\sqrt{n}})$ .

PROOF. (i) follows directly from [Corollary 3](#) and [Proposition 11](#).

For (ii), recall the [Wielandt-Hoffman Theorem](#) and note that

$$\begin{aligned}\lambda_m - \lambda_{m+1} - (\sigma_m - \sigma_{m+1}) &\leq \sqrt{2} \|\mathbf{E}Z - \bar{Z}_n\|_F \\ \Leftrightarrow \lambda_m - \lambda_{m+1} &\leq \sigma_m - \sigma_{m+1} + \sqrt{2} \|\mathbf{E}Z - \bar{Z}_n\|_F\end{aligned}$$

where  $\sigma_m$  and  $\sigma_{m+1}$  denote the  $m$ -th and  $m+1$ -th eigenvalues of  $\mathbf{E}Z$ . If  $\pi(\mathbf{E}Z)$  is not unique, then  $\sigma_m = \sigma_{m+1}$ . By [Corollary 3](#),

$$\|\mathbf{E}Z - \bar{Z}_n\|_F < \sqrt{\frac{1}{\alpha n} \left(1 - \|\bar{Z}_n\|_F^2 + \frac{1}{\alpha n}\right)}$$

with probability at least  $(1 - \alpha)$ .

For (iii),

$$\begin{aligned}\mathbf{P} \left( \lambda_m - \lambda_{m+1} \geq \sqrt{\frac{2(1 - \|\bar{Z}_n\|_F^2 + \frac{1}{\alpha n})}{\alpha n}} \right) &\geq \mathbf{P} \left( \sigma_m - \sigma_{m+1} - \sqrt{2} \|\mathbf{E}Z - \bar{Z}_n\|_F \geq \sqrt{\frac{2(1 + \frac{1}{\alpha n})}{\alpha n}} \right) \\ &= \mathbf{P} \left( \|\mathbf{E}Z - \bar{Z}_n\|_F \leq \frac{\sigma_m - \sigma_{m+1}}{\sqrt{2}} - \sqrt{\frac{1 + \frac{1}{\alpha n}}{\alpha n}} \right) \\ &\geq 1 - e^{-\frac{1}{2} \left( \sqrt{n} \frac{\sigma_m - \sigma_{m+1}}{\sqrt{2}} - \sqrt{\frac{1 + \frac{1}{\alpha n}}{\alpha}} - \sqrt{1 - \|\mathbf{E}Z\|^2} \right)^2} \\ &\rightarrow 1,\end{aligned}$$

using results of ([Boucheron et al.; 2013](#)). □

## 5 Application to shape spaces

Consider the projective shape space  $P\Sigma_m^k$ ,  $k > m + 2$  consisting of all shapes with a projective frame in the first  $m + 2$  points. A result of [Mardia and Patrangenaru \(2005\)](#) is a diffeomorphism  $P\Sigma_m^k \cong (\mathbb{R}P^m)^q$  with  $q = k - m - 2$ , hence non-asymptotic confidence sets can be computed using [Proposition 12](#). For the sake of readability, let  $q = 1$  throughout this example. Recall that the Euclidean space  $\mathbb{R}^m$  can be embedded in  $\mathbb{R}P^m$  preserving collinearity, e.g. by using homogeneous coordinates

$$x = (x_1, \dots, x_m)^t \mapsto [x : 1] := [x_1 : \dots : x_m : 1]$$

missing only the hyperplane at infinity. Using this embedding, a point  $[X_1 : \dots : X_m : X_{m+1}] \in \mathbb{R}P^m$  has a representative in  $\mathbb{R}^m$  if and only if  $X_{m+1} \neq 0$ . Given data in  $P\Sigma_m^k$ , we want to illustrate a confidence set for the extrinsic mean in  $\mathbb{R}^m$  if the extrinsic sample mean  $\mu$  is not in the infinity hyperplane, i.e. has a representative in  $\mathbb{R}^m$ .

Define  $\tilde{x} := (x^t, 1)^t \in \mathbb{R}^{m+1}$  for  $x \in \mathbb{R}^m$  and let  $[z : 1] = \mu$ . Then for all  $x \in \mathbb{R}^m$  with

$$d^2([x : 1], [z : 1]) = 2 - \frac{2}{(\|x\|_2^2 + 1)(\|z\|_2^2 + 1)} (x^t z + 1)^2 < \delta_n^2$$

the following holds:

$$(2 - \delta_n^2)(\|z\|_2^2 + 1) < 2(x^t z + 1)^2 - (2 - \delta_n^2)\|x\|_2^2(\|z\|_2^2 + 1).$$

Hence, visualizing confidence sets for extrinsic means in  $\Sigma_m^k$  is understanding quadrics in  $\mathbb{R}^m$ . Particularly in the case  $m = 2$ , these quadrics represent cone sections such that the confidence sets are the “interior” of these.

A synthesized data set will be used to show the value of this discussion. Consider  $n = 100$  points  $(10 + x_1, 10 + x_2) \in \mathbb{R}^2$  with  $x_1, x_2$  being random real numbers generated by  $\mathbf{R}$  with univariate distribution (mean 0, standard derivation 1). Embedding of  $\mathbb{R}^2$  in  $\mathbb{R}P^2$  produces data in  $\mathbb{R}P^2$  resp. in  $\text{Herm}(3)$ . The resulting Euclidean sample mean is

$$\begin{pmatrix} 0.5075 & 0.4917 & 0.0503 \\ 0.4917 & 0.4875 & 0.0492 \\ 0.0503 & 0.0492 & 0.005 \end{pmatrix},$$



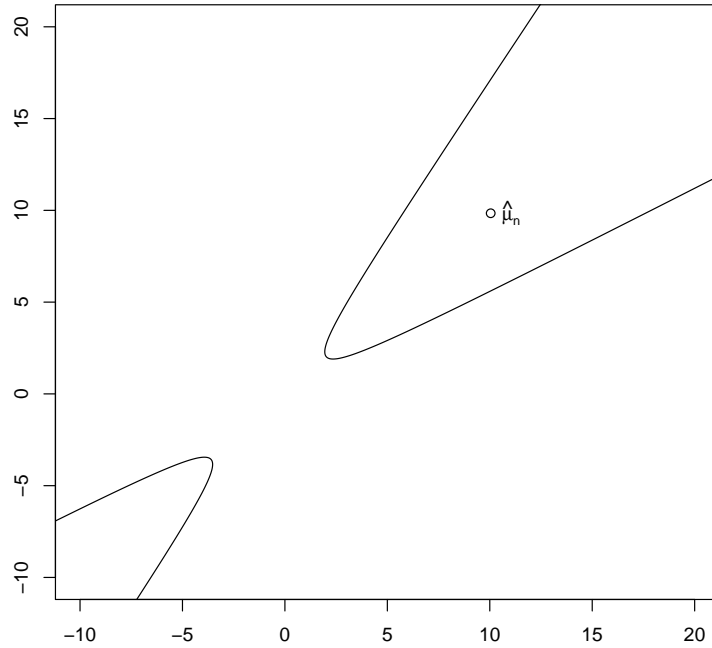


Figure 2: Here, the confidence set to the synthesized data from [Section 5](#) is plotted. This confidence set is a subset of  $\mathbb{R}^2$  when inhomogeneous coordinates are used. The confidence set contains the extrinsic sample mean  $\hat{\mu}_n$ , and the boundary of the confidence region is a hyperbola.

hence the extrinsic sample mean is

$$\begin{bmatrix} 10.0447 \\ 9.8422 \\ 1 \end{bmatrix}$$

and  $\delta_{100} = 0.3713$ . See [Figure 2](#) for the visualization.

## 6 Discussion and outlook

We showed how to construct non-asymptotic and rate-optimal confidence sets for the extrinsic population mean for i.i.d. data on spheres resp. Grassmann manifold. Unfortunately, these are too big in comparison to asymptotic confidence sets to be of practical use due to the fact that the Chebychev inequality has rather loose bounds and is not sharp for bounded random variables. Therefore, sharper inequalities for multivariate, bounded random variables would result in smaller confidence regions. Additionally, one would like to take the sample covariance into account as in ([Hotz et al.; 2015](#)). Unfortunately, the construction of the confidence sets gets tougher for the known multivariate mass concentration inequalities whence the construction of these is one of the aims of future research.

## References

- Bhattacharya, R. and Patrangenaru, V. (2003). Large sample theory of intrinsic and extrinsic sample means on manifolds. i, *The Annals of Statistics* **31**(1): 1–29.
- Boucheron, S., Lugosi, G. and Massart, P. (2013). *Concentration Inequalities : A Nonasymptotic Theory of Independence*, Oxford University Press, Oxford.
- Davis, C. and Kahan, W. M. (1969). Some new bounds on perturbation of subspaces, *Bulletin of the American Mathematical Society* **75**(4): 863–868.

- Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié, *Annales de l'I.H.P.* **10**(4): 215–310.
- Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *Journal of the American Statistical Association* **58**(301): 13–30.
- Hotz, T. (2013). Extrinsic vs intrinsic means on the circle, *Lecture Notes in Computer Science* **8085**(2013): 433–440.
- Hotz, T., Kelma, F. and Wieditz, J. (2015). Universal, non-asymptotic confidence sets for circular means, in F. Nielsen and F. Barbaresco (eds), *Geometric Science of Information*, pp. 635–642.
- Jammalamadaka, S. and SenGupta, A. (2001). *Topics in Circular Statistics*, Vol. 5 of *Series on Multivariate Analysis*, World Scientific, Singapore.
- Kendall, D. G., Barden, D., Carne, T. K. and Le, H. (1999). *Shape and Shape Theory*, Wiley series in probability and statistics, John Wiley & Sons Ltd, Chichester.
- Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*, John Wiley & Sons Ltd, Chichester.
- Mardia, K. V. and Patrangenaru, V. (2005). Directions and projective shapes, *The Annals of Statistics* **33**(4): 1666–1699.
- Patrangenaru, V. and Mardia, K. V. (2003). Affine shape analysis and image analysis, in R. G. Aykroyd, K. V. Mardia and M. J. Langdon (eds), *Stochastic Geometry, Biological Structure and Images*, pp. 57–62.
- Wielandt, H. W. and Hoffman, A. J. (1953). The variation of the spectrum of a normal matrix, *Duke Mathematical Journal* **20**(1): 37–39.